DOCUMENT RESUME

ED 452 515 CS 014 379

AUTHOR Paris, Alison H.; Paris, Scott G.
TITLE Children's Comprehension of Narrative Picture Books. CIERA Report.
INSTITUTION Center for the Improvement of Early Reading Achievement, Ann Arbor, MI.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
REPORT NO CIERA-3-012
PUB DATE 2001-05-15
NOTE 37p.
CONTRACT R305R70004
AVAILABLE FROM CIERA/University of Michigan, 610 E. University Ave., Room 1600 SEB, Ann Arbor, MI 48109-1259. Web site: http://www.ciera.org.
PUB TYPE Reports - Research (143)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Beginning Reading; *Comprehension; Developmental Stages; *Picture Books; Primary Education; *Reading Achievement; Reading Instruction; *Reading Skills

ABSTRACT

This paper explains the creation and validation of the Narrative Comprehension of Picture Books task (NC task), an assessment of young children's comprehension of wordless picture books. Study 1 explored developmental improvements in the task, as well as relationships to other measures of early reading. Subjects, 158 K-2 students, were administered the NC task. Those children who could read were also given a measure of oral reading and comprehension, while nonreaders received an assessment that measured early literacy skills. There was significant improvement with increasing age on NC task measures. Significant relationships were also found between the NC task and several early literacy skills for nonreaders. Study 2 tested the generalizability of the NC task. A subsample of 91 students received 2 additional picture books. Inter-task correlations showed that children were consistent on each of the NC task-dependent variables across the three books. The same developmental trends by grade and reading ability were evident on all three versions of the task. The NC task appears to be a valid quantitative measure of young children's comprehension that is sensitive to developmental changes and adaptable to other books. The authors discuss how narrative comprehension is fundamental to beginning reading and how the NC task may be used for classroom instruction and assessment. (Contains 20 references and 8 tables of data. Appendixes contain an NC task "picture walk"; comprehension questions; scoring rubrics; a description of the Qualitative Reading Inventory--II and the Michigan Literacy Progress Profile.) (RS)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 452 515

Children's Comprehension of Narrative Picture Books

Alison H. Paris
Scott G. Paris
University of Michigan

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

- ☐ This document has been reproduced as received from the person or organization originating it.
- ☐ Minor changes have been made to improve reproduction quality.
- • Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

CIERA REPORT #3-012
May 15, 2001

CIERA
Center for the Improvement of Early Reading Achievement

BEST COPY AVAILABLE

CS 014 379

Children's Comprehension of Narrative Picture Books

CIERA REPORT #3-012

Alison H. Paris and Scott G. Paris
University of Michigan

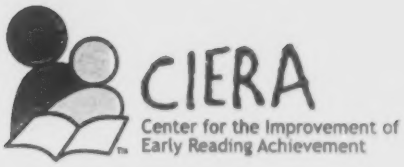
CIERA Inquiry 3: Policy and Profession
How can we assess comprehension when children have limited decoding skills?

IThis paper explains the creation and validation of the Narrative Comprehension of Picture Books task (NC task), an assessment of young children's comprehension of wordless picture books. Study 1 explored developmental improvements in the task, as well as relationships to other measures of early reading. 158 K-2 students were administered the NC task. Those children who could read were also given a measure of oral reading and comprehension, while nonreaders received an assessment that measured early literacy skills. There was significant improvement with increasing age on NC task measures. Significant relationships were also found between the NC task and the comprehension measure for readers, and between the NC task and several early literacy skills for nonreaders. Study 2 tested the generalizability of the NC task. A subsample of students (n = 91) received two additional picture books. Inter-task correlations showed that children were consistent on each of the NC task-dependent variables across the three books. The same developmental trends by grade and reading ability were evident on all three versions of the task. The NC task appears to be a valid quantitative measure of young children's comprehension that is sensitive to developmental changes and adaptable to other books. We discuss how narrative comprehension is fundamental to beginning reading and how the NC task may be used for classroom instruction and assessment.

CIERA

Center for the Improvement of Early Reading Achievement
University of Michigan – Ann Arbor

May 15, 2001



University of Michigan School of Education

610 E University Ave., Rm 1600 SEB
Ann Arbor, MI 48109-1259

734.647.6940 voice
734.615.4858 fax
ciera@umich.edu

www.ciera.org

©2001 Center for the Improvement of Early Reading Achievement.

This research was supported under the Educational Research and Development Centers Program, PR/Award Number R305R70004, as administered by the Office of Educational Research and Improvement, U.S. Department of Education. However, the comments do not necessarily represent the positions or policies of the National Institute of Student Achievement, Curriculum, and Assessment or the National Institute on Early Childhood Development, or the U.S. Department of Education, and you should not assume endorsement by the Federal Government.

Children's Comprehension of Narrative Picture Books

Alison H. Paris and Scott G. Paris
University of Michigan

One of the fundamental controversies regarding children's beginning reading is the relative importance of bottom-up (e.g., decoding) versus top-down (e.g., comprehension) processes. Arguments over whether children use decoding to construct meaning or use meaning to guide decoding underlie historical controversies such as "Why Johnny Can't Read" (Flesch, 1955), the "great debate" (Chall, 1967), the "reading wars" (Edmondson, 1998), and most recently, political legislation regarding phonics instruction (The Reading Excellence Act, HR2614, 1998). We find it paradoxical that, in all these historical debates, less attention has been given to analyses of top-down processes compared to bottom-up processes. For example, in the influential book *Beginning to Read* by Marilyn Adams (1991), scant attention is given to young children's meaning-making. Adams relegates comprehension to a mysterious "meaning processor" that interacts with a "context processor" in an information-processing model. In contrast, most of the book's other chapters provide detailed evidence about the importance of phonological skills, phonics instruction, and phonemic awareness for reading words.

The list of top-down processes for beginning readers usually includes language development, memory skills, vocabulary, and concept development, all of which are related to intellectual development and perhaps experiential factors in early childhood, but not specifically to reading. Perhaps this is why Whitehurst and Lonigan (1998) refer to two classes of emergent literacy as "outside-in" and "inside-out" processes. The outside-in processes include language, narrative, conventions of print, and emergent (i.e., pretend) reading. All of these processes influence beginning reading but are also interconnected with other domains of development. In contrast, inside-out processes, which include knowledge of graphemes, phonological awareness, syntactic awareness, phoneme-grapheme correspondence, and emergent writing (i.e., phonetic spelling), are specifically related to cracking the letter-sound code. Whether we refer to the processes as top-down or outside-in, it is clear that they are general aspects of cognitive and linguistic development that are not specified in the same detail as bottom-up or inside-out processes. Even advocates of "whole language" who champion top-down positions (e.g., Goodman, 1986) have not defined specific processes and devised detailed measures to the same extent as bottom-up researchers.

The difference in specificity between the general classes of processes that influence reading acquisition illuminate two other paradoxes about early reading. First, few measures of outside-in processes are specifically related to

reading. The best, and most widely used, of these is probably Clay's (1985) Concepts About Print (CAP) task, which Stallman and Pearson (1990) praised as an assessment tool because it engages children with actual books in authentic reading situations. However, the CAP assesses isolated aspects of early reading knowledge such as directionality of print and the meaning of punctuation marks, which vary in importance for children learning to read. Other outside-in assessments include language measures such as the PPVT, which may be more closely related to intellectual development than to specific reading skills. Although Whitehurst and Lonigan (1998) do not include perceptual and memory processes as outside-in factors, others might, but those factors are still general developmental accomplishments rather than skills specific to reading. The paradox is that inside-out processes such as phonological awareness are specified in great detail, are closely connected to beginning reading, and have multiple assessment tasks for researchers and educators, in contrast to outside-in processes, which do not share the same consensus about definition or sensitivity of measurement tools. Perhaps this is part of the reason why the great debate is currently being won by advocates of the bottom-up position. It may also explain in part why Whitehurst and Lonigan (1998) observed that inside-out skills predicted reading in first grade better than outside-in skills. More directly, conceptual claims about these contrasting positions are confounded by methodological differences.

A second paradox in the relationship between inside-out and outside-in processes is the imbalance in beginning readers' everyday reading practices. We believe that the usual joint reading practices of parents and children prioritize meaning-making and outside-in processes. This is evident from parents' conversations with their children and the questions they pose to them. It is evident in the use of familiar materials such as alphabet books, counting books, picture books of animals and objects, and similar materials that foster schema-driven comprehension. And it is evident in the repeated reading experiences that transfer recitation, interpretation, and questioning of text into child-controlled processes. Creating, discovering, and sharing meaning are fundamental outside-in processes that bridge oral language to text processing (Burns, Griffin, & Snow, 1999). They provide purpose and structure to early reading, and they motivate children to stay engaged with books and expend effort at toward cracking the code. It seems paradoxical to us that beginning reading instruction in school should contradict early reading practices found at home. Code-based instruction that minimizes the materials, purposes, cognitive processes, and social experiences that motivate children's desire to read at home may dramatically alter the ways that children approach reading.

These three paradoxical imbalances, i.e., the differences in specificity of skills, availability of assessment tools, and emphases at school and home, influence the "balancing" of bottom-up and top-down processes and their relative importance for learning to read. The present research redresses these imbalances somewhat through the creation of a measure of young children's comprehension skills. We have developed a meaning-making task that is connected to reading text by using wordless picture books to assess children's narrative comprehension. The skill reflects the ability to "read" and understand pictures, connect them into meaningful sequences, and make inferences about relationships within and between pictures. These picture books are not just sequences of isolated pictures, but rather stories that have

structural narrative elements including settings, characters, plots, and resolutions. The interpretive and constructive skills involved in narrative comprehension of picture books are parallel to the text-processing skills involved in reading narratives, but do not involve the decoding of printed words (e.g., Yussen & Ozcan, 1996). Assessment of children's narrative comprehension is thus an outside-in process that is specific and connected to reading. It is also "authentic," because it is consistent with shared reading experiences of children and adults who examine and interpret pictures, whether text is present or not, in order to understand the story.

Narrative comprehension is also a novel assessment task that may permit measurement of the cognitive processes that directly contribute to early reading. It is one example of a schema-driven comprehension approach; we can imagine other, similar assessment tasks for young children based on schemata of numbers, the alphabet, repetitive rhymes, or expository genres. We have chosen narrative because of the primacy of the role it plays in people's thinking (Bruner, 1986) and in children's early interactions with text. Our definition of reading extends beyond simply saying the words on the page to include understanding their literal and implied meanings. Likewise, it goes beyond identification of individual pictures to look at narrative comprehension of pictorial sequences and stories. Whitehurst and Lonigan (1998) described narrative as one of four outside-in processes and suggested that research, assessment, and instructional attention has not focused enough on the process of "understanding and producing narrative" (pg. 850) as a component of early literacy.

There are limited instruments available for assessing picture book comprehension, particularly at the narrative level. Morrow (1990) assessed children's "construction of narrative" by observing child-adult discussions while reading, by eliciting a retelling, and by categorizing the child's behaviors according to where the focus was placed (e.g., on print, story structure, illustrations, print, questions, comments). Although Morrow identifies this process as a measure of narrative meaning-making, it is really more about the social process of meaning-making and the interactions that occur among adult, child, and book. Her measure does not necessarily represent the quality of children's comprehension and certainly is not specific to narrative-level thinking which requires integration of information. Similarly, Sulzby (1985) described how children's language displayed a developmental progression, from viewing individual storybook pages as if they were discrete units to treating the pictures as parts of an integrated whole. Although Sulzby differentiates between page- and story-level descriptions, her assessment is actually a measure of the development of speech patterns rather than comprehension of specific story information.

Van Kraayenoord and Paris (1996) created an assessment called "Story Construction from a Picture Book," which measured Australian children's abilities to construct meaning from pictures. Six aspects of meaning-making were measured, including initial examination of the book, remarks about the pictures, elaboration, metalinguistics, revision strategies, and identification of themes or morals. The authors found that children's abilities to construct stories from picture books at 5 to 6 years of age were correlated significantly with standardized reading test scores two years later.

The present study aims to rectify the lack of attention and tools for assessing children's narrative-level comprehension. It is a revision of the task created

by van Kraayenoord and Paris (1996) that extends their assessment by using better materials, more subjects, a wider range of ages and reading abilities, and by examining correlations with a greater variety of reading abilities. The primary purpose of the new assessment is to provide a measure of children's comprehension of narrative events independent of their ability to decode and read words in a narrative story. We believe that children's narrative thinking is a fundamental contributor to early reading comprehension and that assessments of narrative thinking with pictures can identify children with comprehension difficulties (or strengths) and suggest areas of instruction for them that are based on reasoning about text messages. Such a comprehension focus in early reading assessment and instruction complements the traditional focus on basic text-decoding skills. The two studies in this paper describe the assessment tool and provide evidence about reliability, generalizability, and developmental patterns of performance seen on the narrative comprehension task.

Study 1

We discovered that many picture books for young children do not provide stories in a narrative genre. Instead, they show pictures related to themes (such as animals or transportation) or illustrate sequences (such as numbers) or provide predictable schemata (such as *Goodnight Moon*) as ways to connect outside-in processes to book comprehension. Text often accompanies narratives for children in familiar folktales or novel stories and it is customary for adults to read the text and children to read the pictures. In order to remove the decoding demands on children and still provide a coherent narrative story, we needed to create a narrative picture book without text. Van Kraayenoord and Paris (1996) removed the supporting text from a trade book—a strategy which can be used by others for both assessment and instructional purposes. For the purposes of the current study, we located commercially published wordless picture books and adapted them by deleting some irrelevant pages to shorten the task, thus creating the Narrative Comprehension (NC) task.

The first study reports our procedures for observing how children interact with wordless picture books under three conditions: spontaneous examination during a "picture walk," elicited retelling, and prompted comprehension during questioning. The picture walk procedure allows children to become familiar with the story before being questioned, a practice recommended for work with young children (Fountas & Pinnell, 1996). The retelling phase provides a measure of the child's free recall of the story. The prompted comprehension phase provides a uniform and quantified procedure for eliciting and scoring children's understanding of narrative elements and relationships. We administered the NC task to children in grades K–2 in order to assess the developmental appropriateness of the testing procedures and the measurement sensitivity of the task.

We also collected additional data about the children's reading performance so that we could determine how performance on the NC task is correlated with other emerging reading skills. Our chosen reading task was the Qualita-

tive Reading Inventory II, QRI-II (Leslie & Caldwell, 1995). The QRI-II is an informal reading inventory that is administered to children individually and provides diagnostic information about children's (a) reading words in isolation; (b) oral reading accuracy; and (c) comprehension and memory of text. The task includes use of narrative and expository passages arranged by grade level and difficulty. In addition, we collected data from several tasks included in the Michigan Literacy Progress Profile (MLPP), including (a) phonemic awareness (PA); (b) hearing and recording sounds (HRS); and (c) Concepts About Print (CAP). Thus, we can correlate performance on the NC task with QRI-II measures for children who can read, and with two inside-out processes (PA and HRS) for nonreaders.

Method

Subjects

158 students from one K-2 school in a midwestern city participated in the study during the fall of 1998. Students were randomly selected from among those who returned permission letters in fourteen classrooms. Their ages ranged from 61-99 months (M = 81, SD = 10). There were approximately equal numbers of females and males, and the sample was ethnically diverse, with 49% Caucasian-American, 22% African-American, 12% Asian-American, and 14% other or multi-racial, as shown in Table 1. Almost half of the students were nonreaders at the start of the study.

Table 1: Characteristics of Participants (N = 158)

Variable	Frequency	Percent of Total
Gender		
Female	81	51.3
Male	77	48.7
Grade		
Kindergarten	34	21.5
First	61	38.6
Second	63	39.9
Ethnicity		
African American	36	22.1
Asian American	19	11.7
Caucasian	80	49.1
Other	22	13.5
Reading Status		
Nonreader	73	46.2
Reader	85	53.8

Task

The book used for this assessment was *Robot-Bot-Bot* by Fernando Krahn, which tells the story of a family whose new robot "housecleaner" goes wild after the child plays with its wires. The book uses black line drawings with no accompanying text and has a clear story line with an obvious sequence of

events and main elements of stories (i.e., settings, characters, problems, resolutions). The adapted version of the robot book omitted a few pictures. The remaining 18 pages were photocopied and assembled into a book format with a spiral binding and cover. The title and author's name on the front cover were the only words in the book.

The task has three parts (Picture Walk; Retelling; Prompted Comprehension) which yield five composite scores: (1) spontaneous reactions to the story; (2) retelling of the story; comprehension of (3) explicit and (4) implicit story information; and (5) total storybook comprehension.

Part 1: Storybook Picture Walk. The Picture Walk uses observations to capture children's spontaneous and independent interactions while "reading" the picture book. The children were first given a closed book and asked to look through it. They were then encouraged to "say out loud whatever you are thinking about the pictures or the story." While the child "read" the story, observations were made about behaviors according to an observational scheme. The scheme specified the following five types of behaviors, or "Picture Walk Elements": Book Handling Skills; Engagement Behaviors; Picture Comments; Storytelling Comments; and Comprehension Strategies (see Appendix A for a complete description).

Book Handling Skills were defined to include how the child orients the book, whether (s)he has a sense of appropriate speed and order, and whether pages are skipped or skimmed. *Engagement* was defined as behavioral and emotional involvement, as judged by attention, interest, affect, and effort. *Picture Comments* described whether the child made discrete comments about a picture, including descriptions of objects, characters, emotions, actions, or opinions, as well as character vocalizations. *Storytelling Comments* were defined as integrative comments across pictures, which demonstrated an understanding that the pictures tell a coherent story. This might include narration, dialogue, and the use of storytelling voice or language. *Comprehension Strategies* evaluated whether the child displayed vocalizations or behaviors which showed attempts at comprehension, such as self-correction of story elements or narrative; looking back or ahead in the book in order to aid in creating the narrative; asking questions for understanding, and making predictions about the story.

In a right-hand column adjacent to each of these elements, a 0–2 point scoring rubric described the behaviors appropriate for a 0-, 1-, or 2-point score, and represented the depth of spontaneous reactions. For example, a child received 0 points for Storytelling Comments if (s)he gave no verbalizations about the pictures, 1 point for inconsistent and discrete provision of storytelling elements, and 2 points for storytelling comments that connected the events of the story through dialogue or narration. The experimenter made the 0–2 point judgments on each element as the child engaged in the Picture Walk, and children received a total Picture Walk Score that could range from 0 to 10, with higher scores indicating higher levels of spontaneous interaction with the book.

Part 2: Retelling. Immediately following the Picture Walk, the book was taken away from the child, and the child was asked to retell as much of the story as possible. When the child completed the retelling, one prompt was given by asking the child if (s)he could remember anything else about the

story. Children's retellings were transcribed, and the information was categorized according to the six following story grammar elements: setting, characters, goal/initiating event, problem/episodes, solution, and resolution/ending. One point was awarded for phrases indicating the presence of each story element. Retelling scores ranged from 0 to 6, with 0 points demonstrating that the child did not recall any of the story elements and 6 signifying that the child's retelling included phrases representing all of the elements.

After completing the retelling and prior to beginning part 3, children were asked two "Previous Experience" questions: whether they had read the robot picture book before, and whether they had read a book similar to this one before. Both responses were coded for yes or no answers.

Part 3: Prompted Comprehension. The third part of the NC Task is designed to assess children's level of narrative comprehension, defined as the construction of meaning from pictures by integrating information across pages so as to create coherent and connected understandings. Following the retelling, the child was told that the experimenter and child would go through the story together a second time while the experimenter asked several questions about the pictures. The experimenter guided the page-turning and elicitation of reactions during this viewing of the book by pointing out pictures and asking a series of ten comprehension questions. Five of these questions were about explicit information and required the identification of the following story elements: characters, setting, initiating event, problem, and outcome resolution. Discussion of the latter three elements was followed by "why" questions in order to promote responses that demonstrated narrative comprehension.

The remaining five comprehension questions required children to make inferences from the pictures about the characters' feelings, dialogue, causal inferences, predictions, and themes. The questions about implicit information were also followed by "why" probes in order to distinguish between children who could make "shallow" inferences and those who could connect the inferences to other story events. (See Appendix B for Prompted Comprehension Questions).

Regardless of their responses to each question, children were always provided with one prompt (i.e., "Is there any other reason why the characters were feeling that way?"). Both the uniform prompts and the "why" questions provided an impetus for narrative thinking. This procedure ensured that all children were given equal opportunities to respond.

Scoring Children's Responses

A scoring rubric was created for the Prompted Comprehension questions based on the assumption that higher levels of narrative understanding would be demonstrated by the integration of information across pictures rather than a focus on describing a single picture in isolation. This criterion was applied to all ten questions by using a 0-1-2 point rubric. In general, 0 points indicated no answer, irrelevant, wrong or inappropriate answers; 1 point represented appropriate answers derived from single pictures; and 2 points were awarded when information from multiple pictures was used to create a coherent explanation consistent and connected with the child's unfolding narrative. This general framework was utilized for each of the ten questions, which played out differently depending on what "narrative com-

prehension" versus "focusing on single pages" meant for each item. (See Appendix C).

The ten prompted comprehension questions yielded three composite scores: "Explicit Comprehension," "Implicit Comprehension," and "Total Prompted Comprehension." The explicit and implicit comprehension subscores ranged from 0 to 10 points. The total comprehension score included all ten questions and ranged from 0 to 20 points. Higher scores on all three scales represent higher levels of narrative comprehension.

Inter-Rater Reliability

In order to ensure that the prompted comprehension rubrics yielded consistent scores across raters, 30% of the sample was randomly selected for an inter-rater reliability check. There was an equal representation of children by grade and reader/nonreader status in this subsample. Two research assistants were trained to use the rubrics and scored children's responses to comprehension questions. Scores were checked for agreement by individual question and tabulated by item and total percent agreement. Inter-rater reliability was above 90% agreement for every item, with a mean of 97% agreement across all items.

Inter-rater reliability was also checked for the retelling measure. An independent rater was trained in use of the rubric, and 30% of the retellings were randomly selected for scoring by the rater. Scores were checked for agreement by total retelling points, and the percent-match was calculated. Retelling reliability was above 90%. We were not able to perform reliability checks on the Picture Walk, because scoring judgments were made as the researcher watched the child "read" the pictures.

The Qualitative Reading Inventory-II (QRI-II).

We assessed five distinct reading skills with the QRI-II assessment procedures: (a) reading of isolated words in graded word lists; (b) oral reading fluency with miscue analysis; (c) retelling of the passage; (d) answering of comprehension questions; and (e) total reading time. (See Appendix D for a description of QRI-II measures.)

Graded Word Lists. The graded word lists of the QRI-II assessed children's abilities to read isolated words. This skill is assessed by noting children's ability to read lists of 20 words in isolation. For each list, three scores were obtained: percentage of words identified automatically (less than one second); percentage of words identified with analysis (greater than one second); and total percentage of words correctly identified. All participants received two word lists.

Oral Reading. Children's oral reading was assessed according to their performance on two of the graded reading passages. Children who struggled to read the lowest level passage received no passage or one passage, depending on whether they could complete the most elementary story (the preprimer level). The procedures yielded the following dependent variables for oral reading: rate of oral reading, percent accuracy, percent acceptability, and percent of self-corrections. Fluent readers, compared to less fluent readers, would be expected to read text with fewer miscues, fewer meaning-changing miscues, and higher rates of self-corrected miscues.

Retelling. Students were asked to retell as much of the story as possible and were given one prompt for any additional information remembered at

the end of the retelling. Retellings were initially scored according to the propositions supplied in the QRI-II manual, so that students received a "propositions recalled" score indicating the number of propositions that the child could remember from the story. In addition, an alternative recall scoring system based on narrative story structure was created, in which children could receive 0–6 points depending on whether their recall included information about six story elements: setting, characters, initiating event, problem, solution, and resolution/ending.

Comprehension. Comprehension was assessed by scoring children's answers according to the QRI-II manual. There were 5 to 8 questions per passage and each set of questions included both inferential and literal questions.

Michigan Literacy Progress Profile (MLPP)

The MLPP (1998) was developed by the Michigan Department of Education to assess multiple features of children's early literacy and includes a variety of assessments, milestone tasks (oral reading fluency, reading comprehension, writing, oral language, and attitudes and self-perceptions), and enabling skills (CAP, letter sound identification, PA, decodable word lists, known words activity, and HRS). Students were assessed on three tasks—CAP, PA, and HRS—described in Appendix E.

Concepts of Print. The CAP task assesses children's knowledge of the fundamental features of text as they examine printed text. The task contains 22 questions which fall into six main categories: Book Concepts; Reading Concepts; Directionality; Concept of Word; Concept of Letter; and Punctuation Marks. Questions were in a format that resembled the following: "Show me the front of this book" or "Show me with your finger which way I go as I read this page." Students received 1 point for each correctly identified concept, resulting in a CAP scale that ranged from 0 to 22 points, with higher points indicating more knowledge about printed text.

Phonemic Awareness. The PA task assesses the child's understanding of the sound units of language. This task includes three subsections: Rhyming, Phoneme Blending, and Phoneme Segmentation. Children received three subscores ranging from 0 to 8 for each of these sections, in addition to a PA total score which was the sum of these three sections and ranged from 0 to 24 points. Higher scores indicated greater levels of phonemic awareness.

Hearing and Recording Sounds. The HRS task measures children's ability to hear individual phonemes and record them as letters. The task is a two-sentence "story" which children write as the words are heard. Students received one point for each correctly recorded sound; scores could range from 0 to 39, because there were 39 distinct sounds in the two sentences.

Procedure. Participants were individually assessed on the NC Task. A subset of the children were also assessed on either the QRI-II or on the MLPP tasks, depending on whether they had been identified as readers or nonreaders. After building rapport, the child was given the first QRI-II word list, with one word at a time exposed for the child to read. The second list was given in the same fashion. No feedback was provided during the task and all children were praised for their performance. If a child identified fewer than 12 of the 20 words (60%) on the preprimer word list, the student was recorded

as a "nonreader" and the remaining parts of the QRI-II were not administered. All other children were given two passages to read aloud, based on the level of word list they read best. Children were told to read the stories as best they could and that they would then be asked questions about the story. After the oral reading of each passage, retellings were initiated and comprehension questions were asked. The readings, retellings, and comprehension questions were all tape recorded for later scoring. If students who were given the preprimer passage could not decode the first few lines, the QRI-II was terminated and the child was recorded as a nonreader. Hence, participants were identified as nonreaders based on low performance on either the preprimer word list or the preprimer passage. Administering the complete QRI-II took approximately 15 to 20 minutes per child.

After completion of the QRI-II, all students were given the Comprehension of Wordless Picture Books task, using the *Robot-Bot-Bot* picture book. Administering the three parts of the task required approximately 10 to 15 minutes per child. Several children were too shy to respond to any of the Part III comprehension questions and were dropped from the analyses. After completion of the NC Task, the three MLPP activities were administered to all children who had been identified as nonreaders. The MLPP assessments required approximately 15 minutes per child. The tasks were often administered in separate sessions, depending on the attention of the child and the schedule of classroom activities.

Results

Each of the five NC task outcome variables was initially examined in order to verify its distributions. All variables were normally distributed with no ceiling or floor effects. As indicated in Table 2, significant positive correlations between variables were observed. Retelling and Prompted Comprehension scores were more highly correlated with each other than either of them was with Picture Walk.

Table 2: Correlation Matrix for NC Task Measures

VARIABLES	1	2	3	4	5
1. Picture walk	—	.19*	.17*	.21†	.11
2. Retelling		—	.54††††	.53†	.44†
3. Narrative comprehension total			—	.89†	.89†
4. Comprehension explicit subscore				—	.62†
5. Comprehension implicit subscore					—

* $p < .05$
† $p < .01$

Picture Walk Results. Children received scores ranging from 1 to 10, which reflected depth of interaction while independently reading the picture book. Table 3 shows the percent of students scoring 0, 1, and 2 points on each of the Picture Walk behaviors. The percentage of students receiving 2 points for each behavior category decreased as the behaviors became

more complex. Approximately 90% of the participants scored 2 points for book handling skills, whereas only 16% of students received 2 points for comprehension strategies. Picture Walk scores ranged from 2 to 10, with a mean of 7.21 (SD = 2.18). Table 4 shows scores for children as a function of grade level and reading ability. ANOVA on the Picture Walk total scores revealed no significant effects due to grade or reading ability. Because of small sample sizes in two of the cells (one Grade K reader and four Grade 2 nonreaders), the data were recoded so that these particular cells were excluded from ANOVA analyses testing grade-by-reading ability interactions. No interaction effects emerged. Nor did one-way ANOVAs show any gender and ethnicity effects. Children did not score differently depending on whether they said that they had seen the *Robot-Bot-Bot* book previously, but children who had previously read a similar book did score significantly higher than children who responded that they had not read a similar book, $F(1, 156) = 4.71, p < .05$.

Table 3: Percent of Children Receiving Scores of 0–1–2 for Each Picture Walk Behavior

Dependent Variable	0 points	1 point	2 points
Book-handling skills	0.6	9.5	89.9
Engagement	0.6	21.5	77.8
Picture comments	20.3	7.0	72.8
Storytelling comments	19.0	21.5	59.5
Comprehension strategies	54.4	29.7	15.8

Table 4: Developmental Changes in NC Task Measures

NC Task Variables	Grade Means (SD)			Reading Ability Means (SD)	
	K	1	2	Nonreader	Reader
	(N = 34)	(N = 61)	(N = 63)	(N = 73)	(N = 85)
Picture walk	6.9 (2.0)	6.9 (2.3)	7.7 (2.1)	6.9 (1.9)	7.5 (2.4)
Retelling	2.4 (1.81)	3.2 (1.8)	4.3 (1.6)	2.8 (1.9)	4.1 (1.6)
Total comprehension	10.2 (4.34)	12.6 (3.8)	15.5 (2.7)	11.2 (4.0)	15.0 (3.2)
Explicit comprehension	5.1 (2.63)	6.8 (2.1)	8.2 (1.6)	5.9 (2.5)	7.9 (1.8)
Implicit comprehension	5.0 (2.48)	5.8 (2.3)	7.3 (1.7)	5.2 (2.3)	7.1 (1.9)

Retelling Results. Retelling scores ranged from 0 to 6, representing the number of major story elements included in the retelling of the picture story. The overall mean retelling score was 3.47 (SD = 1.85). As indicated in Table 4, mean retelling scores increased significantly with grade level, $F(2, 155) = 15.10, p < .001$. Also, children identified as "readers" scored significantly higher than nonreaders, $F(1, 156) = 22.67, p < .001$. To test for interaction effects, data had to be recoded so that the following cells could be compared: kindergarten nonreaders, first grade nonreaders, first grade readers, and second grade nonreaders. Since there were few kindergarten readers and second grade nonreaders, it was not possible to conduct a 3 (Grade)

x 2 (Reading Ability) ANOVA. Instead, a four-group one-way ANOVA was performed. ANOVA with Scheffe post hoc tests revealed a Grade x Ability interaction, indicating that, on average, second graders performed higher than first grade nonreaders but not higher than first grade readers, $F(3, 149) = 10.56$, $p < .001$. Thus, children who could read text were more successful at retelling stories in pictures. No main effects emerged by gender or ethnicity. Children who had read a similar picture book retold significantly more story elements, $F(1, 156) = 4.55$, $p < .05$.

Prompted Comprehension Results. Children received Total Prompted Comprehension scores ranging from 0 to 20, reflecting the extent to which responses integrated story information across pictures or events. Table 5 shows the percent of students scoring 0, 1, and 2 points on each of the five explicit and implicit comprehension questions.

Table 5: Percent of Children Receiving Scores of 0–1–2 for Each Prompted Comprehension Question

Dependent Variable	0 points	1 point	2 points
Character	6.3	20.3	73.4
Setting	32.3	22.8	44.9
Feelings	13.9	44.9	41.1
Causal inference	12.0	15.2	72.8
Initiating event	10.8	62.0	27.2
Problem	11.4	23.4	65.2
Dialogue	8.9	56.3	34.8
Outcome resolution	4.4	40.5	55.1
Prediction	25.3	37.3	37.3
Theme	38.0	24.7	37.3

It is evident that explicit questions were easier than implicit questions. The most difficult items concerned an appropriate setting, prediction, and theme, whereas easier items queried character, problem, and outcome resolution. Overall means for Total Prompted Comprehension, Explicit Comprehension, and Implicit Comprehension were 13.26 ($SD = 4.05$), 6.98 ($SD = 2.34$), and 6.23 ($SD = 6.23$), respectively. Explicit Comprehension subscores were significantly higher than subscores on Implicit Comprehension, $t(157) = 4.64$, $p < .001$. For all three of these measures, older children scored significantly higher than younger children: Total Comprehension, $F(2, 155) = 26.35$, $p < .001$; Explicit Comprehension, $F(2, 155) = 27.10$, $p < .001$; Implicit Comprehension, $F(2, 155) = 15.37$, $p < .001$.

Scheffe post hoc tests showed that there were significant differences in Prompted Comprehension between successive grades, whereas for Explicit and Implicit Comprehension, significant differences emerged between all grades except kindergarten and first. Paired t-tests were performed in order to examine whether explicit scores were greater than implicit scores at each grade level. The tests revealed that explicit scores were significantly greater than implicit scores for first grade students, $t(60) = 3.87$, $p < .001$, and second grade students, $t(62) = 4.11$, $p < .001$, whereas kindergartners' explicit and implicit scores did not significantly differ. Additionally, readers scored

significantly higher than nonreaders on Total Prompted Comprehension, $F(1, 156) = 43.77$, $p < .001$; Explicit Comprehension, $F(1, 156) = 36.10$, $p < .001$; and Implicit Comprehension, $F(1, 156) = 31.02$, $p < .001$.

After recoding the data to test for interaction effects as in the retelling analyses, ANOVA indicated that whether grade 1 students performed significantly below or above kindergartners or grade 2 students depended on the reading ability of the grade 1 students, $F(3, 149) = 20.09$, $p < .001$. Scheffe post hocs showed that second grade students scored significantly higher than first grade nonreaders but not higher than first grade readers. In results similar to those from the retelling task, older children and children who could read text were better at integrating pictures in order to answer explicit and implicit comprehension questions. No gender or ethnicity main effects or interactions were found.

Relationships with other reading variables for nonreaders. Intercorrelations among variables for nonreaders were examined between NC task outcomes and MLPP measures. MLPP tasks were not significantly correlated with Picture Walk or Retelling measures, but did show strong relationships with Prompted Comprehension. Prompted Comprehension was significantly correlated with Phoneme Segmentation ($r = .35$, $p < .01$), PA Total ($r = .33$, $p < .01$), HRS ($r = .33$, $p < .01$), and CAP ($r = .44$, $p < .01$). In a hierarchical regression on Prompted Comprehension, age was entered in the first step and was a significant contributor to Prompted Comprehension, $b = .16$, $p < .01$; $R2 = .11$, $p < .01$. At step 2, Picture Walk, Retelling, Concepts about Print, and Phonemic Awareness were entered, causing an additional R2 change of .23 ($p < .001$). Retelling and CAP, both outside-in processes, were significant predictors of Prompted Comprehension, $b = .62$, $p < .01$; and $b = .40$, $p < .05$, respectively. Apparently, entering age into the regression equation first nullified any relationship between Phonemic Awareness and Prompted Comprehension. The overall model accounted for 35% of the variance in Prompted Comprehension.

Relationships with other reading variables for readers. Intercorrelations among variables for readers were examined between NC task outcomes and QRI-II retelling and comprehension measures. Neither Picture Walk nor Prompted Comprehension correlated significantly with the QRI-II comprehension or retelling measures. Significant relationships emerged between NC Retelling and QRI-II Comprehension ($r = .27$, $p < .05$) and between NC Retelling and QRI-II Retelling ($r = .26$, $p < .05$).

Discussion

The NC task appears to be a useful quantitative measure of young children's narrative comprehension. It assesses young children's thinking and comprehension of narrative sequences without the conflation of decoding skills. Hence, children's early and emerging comprehension can be evaluated even though they may not be able to decode text. The Picture Walk, Retelling, and Prompted Comprehension measures had normal distributions without floor or ceiling effects, and the procedures were appropriate for four- to eight-year-old children whether they could read or not. The task indicates a developmental progression in Retelling and Prompted Comprehension but shows

no age-related differences in children's spontaneous examinations of picture books. Similarly, reading ability differences emerged from the NC Task, with readers performing better than nonreaders on retelling and comprehension. There were no differences by gender or ethnicity on any of the measures.

The correlations of several reading skills were examined for nonreaders. Phonemic awareness and CAP were correlated significantly only for Prompted Comprehension. This suggests that skills involved in looking at pictures during the Picture Walk and skills involved in oral retelling were not linked to specific metacognitive awareness about language sounds or text features. Considering these findings together with the above developmental trends, it appears that the Picture Walk has less developmental sensitivity, implying that four- to eight-year-olds look at picture books similarly and that what they do may not be related to age. Additionally, the strong relationship between Retelling and Prompted Comprehension suggests that the two outside-in processes are strongly correlated skills that underlie narrative comprehension. Moreover, as shown by the regression, Phonemic Awareness, an inside-out process, does not seem to be measuring the same skills as Prompted Comprehension and Retelling, which leads to the possible conclusion that narrative comprehension may in fact be an outside-in process and an independent factor of early literacy.

Regarding the NC task and QRI-II retelling and comprehension measures, the lack of relationship between Prompted Comprehension and QRI-II comprehension may indicate the different ways that comprehension is assessed in the two tasks. Prompted Comprehension was specifically coded for differentiating identifying elements and explaining them at the narrative level, whereas the comprehension measure on the QRI-II awards equivalent points regardless of the complexity of thinking. However, the significant relationship between NC task retelling and QRI-II retelling helps to validate the use of "retelling" in the NC task as good evidence about story understanding.

The NC task has the positive properties of assessment instruments identified by Stallman and Pearson (1990). It assesses reading while children are engaged with an authentic book and is consistent with the types of interactions that children have with parents around picture books. Additionally, the task provides both consequential and curricular validity (Linn et al., 1991) because the task can be aligned with instructional practices and can have positive consequences for children if it is used to help create narrative-level meanings from picture books. It can also be used by elementary school teachers and intervention programs for at-risk kids as an instructional tool that strengthens outside-in processes for beginning reading. For example, it makes clear the importance of encouraging comprehension strategies and eliciting storytelling comments, and it outlines the story elements that young readers should be able to identify and connect to the entire narrative.

As stated by Snow and Ninio (1986), "books are a source of enchantment and wonder. This might, after all, turn out to be the most important contribution of picture-book reading to the acquisition of literacy." If personal and narrative meaning-making contribute to children's enchantment with reading, it is important to provide instruction that fosters this wonder and to assess whether children are engaging in the kinds of activities which have the power to draw them forever into the world of books and literacy.

Study 2

If narrative competence is a general characteristic of children's thinking and reading, it should be evident in a variety of materials and stories. Van Kraayenord and Paris (1996) and Study 1 used different picture books and found similar results, but a test of the generalizability of narrative competence was still needed. The purpose of Study 2 was to examine the reliability and generalizability of the NC task across three picture books. Because the task was created initially using a specific book (*Robot-Bot-Bot*), it was necessary to determine whether the protocol for task administration, the types of questions asked, and the scoring rubrics could be applied to other picture books in a manner comparable to the original NC task. Thus, Study 2 tested the generalizability of the NC task and the reliability of the developmental performance patterns observed in Study 1.

Method

Subjects

Study 2 participants included a subsample of Study 1 nonreaders (excluding three students who did not complete all three storybooks) and a new, randomly selected sample of readers (n = 91). The students in Study 2 were in kindergarten (n = 31), grade 1 (n = 46), and grade 2 (n = 14); their ages ranged from 61 to 98 months (M = 77, SD = 9). There were approximately equal numbers of females and males, and the sample was ethnically diverse: 58.2% Caucasian, 18.7% African American, and 6.6% Asian American.

Task. Two additional picture books were selected to test the NC task: Mercer Mayer's *A Boy, A Dog, and A Frog* and Fernando Krahn's *The Magic Carpet*. In the "frog" book, a boy leaves a pond after unsuccessfully trying to catch a frog, who then becomes saddened by the boy's departure. The "carpet" book tells the story of a girl who receives a magic carpet that flies out of control. Similar to the robot book, both the frog and the carpet books used black line drawings with no accompanying words and had clear story lines with an obvious sequence of events and narrative structure. Several pictures were omitted in order to create the adapted versions of these books. The remaining twenty-four pages of the frog book and twenty-six pages of the carpet book were photocopied and assembled into book format with spiral bindings and covers. The only words in the adapted book versions were the title and authors' names on the front cover. For reference purposes, the three versions of the task will be identified as NC task-R (the robot book), NC task-F (the frog book), and NC task-C (the carpet book).

The same three-part format (Picture Walk, Retelling, Prompted Comprehension) was applied to the new books. The Picture Walk did not need to be modified for Study 2 books because the original observation scheme and scoring system can be applied to children's examination of any book. Retelling was also scored according to the same six narrative elements (setting, characters, goal/initiating event, problem/episodes, solution, and resolution/ending) but was identified for information specific to the frog and carpet books. The order and spacing of questions for Prompted Comprehension were modified for NC task-F and NC task-C. The same five explicit and

implicit comprehension questions were used, but the order had to be changed so that the questions occurred on pages appropriate to the stories.

The only question that had to be completely changed for the modified versions of the task was the final "Theme" question, since that question was created specifically for each book. For NC task-F the theme question was: "In thinking about everything that you learned after reading this book, what would you tell a boy who was going out to a pond to catch a frog in order to help him? Why would you tell him that?" The theme question in NC task-C was, "In thinking about everything that you learned after reading this book, if your friend had a magic carpet that started to fly out of control, what would you say to help your friend so that what happened in this story doesn't happen to him/her? Why would you tell him/her that?" As in NC task-R, appropriate questions were followed by explanatory probes (e.g., "Why do you think so?"), and children's responses were always followed up with one prompt for additional information. The same scoring rubric was used to score the prompted comprehension questions for NC task-F and NC task-C. Higher levels of narrative understanding are represented by the integration of information across pictures rather than a focus on describing a single picture in isolation.

Consistent with the inter-rater reliability checks performed on the Prompted Comprehension questions for NC task-R, 30% of the sample with equal grade and reading ability distributions was randomly selected to score prompted comprehension questions for NC task-F and NC task-C. Two trained researchers scored the ten comprehension questions, and scores were checked for agreement by individual question type and for total percent agreement. For both the frog and carpet stories, percent agreement by item ranged from 89% to 100%, with a 94% average percent agreement across all items. The rubric, therefore, was useful across picture books, allowing raters to make reliable judgments about the degree of narrative-level thinking represented in children's responses to the prompted comprehension questions.

Procedure

The procedure was the same as in Study 1, with the exception that participants in the Study 2 subsample received three picture books rather than just the single robot picture book. After children were given the QRI-II word lists (and reading passages if they were readers), children who had been selected for Study 2 received the NC task-R, NC task-F, and NC task-C. In order to control for order effects, the order of picture book administration was randomly varied. As in Study 1, children first engaged in the Picture Walk, then gave a retelling, and finally were asked the series of prompted comprehension questions. For some children, only one or two books were given in a single sitting; examiners then returned later that day or on the following day to administer the remaining books.

Results

The five NC task measures for each of the three versions were initially examined in order to compare them across picture books. Normal distributions with no ceiling or floor effects emerged for all three versions. Moreover, as shown in Table 6, the means and standard deviations for each of the mea-

sures were very similar, regardless of version. The three books also yielded similar trends by grade and reading ability, as shown in Tables 6 and 7.

Intra-task correlations were then investigated in order to assess whether the NC task outcomes showed similar relationships to each other for the three versions of the task. The measures related to each other in expected directions. With the exception of the correlations between the NC task-R Picture Walk and Comprehension variables, significant positive relationships existed between all variables. Particularly important were the consistent, strong relationships between retelling and Prompted Comprehension, shown in Table 8, which ranged from r = .46 to r = .61.

Table 6: Developmental Changes in NC Task Measures for NC Task-R, NC Task-F, and NC Task-C

NC Task Variables	Overall Means (SD)	Means by Grade (SD)		
		K (N = 31)	1 (N = 46)	2 (N = 14)
Picture walk				
NC task-R	7.14 (1.98)	6.84 (2.05)	7.17 (1.98)	7.71 (1.77)
NC task-F	7.27 (2.20)	6.70 (2.45)	7.28 (2.17)	8.43 (1.09)
NC task-C	7.08 (2.25)	6.94 (2.21)	6.89 (2.48)	8.00 (1.18)
Retelling				
NC task-R	3.05 (1.89)	2.42 (1.88)	3.04 (1.81)	4.50 (1.40)
NC task-F	3.00 (1.77)	2.13 (1.43)	3.17 (1.76)	4.29 (1.64)
NC task-C	2.99 (1.86)	2.16 (1.66)	3.07 (1.74)	4.57 (1.70)
Total comprehension				
NC task-R	12.18 (4.09)	10.35 (4.42)	12.59 (3.51)	14.86 (3.46)
NC task-F	11.88 (3.49)	9.87 (3.09)	12.70 (3.39)	13.50 (2.82)
NC task-C	11.35 (4.35)	9.06 (4.20)	12.00 (3.84)	14.29 (4.01)
Explicit comprehension				
NC task-R	6.38 (2.47)	5.10 (2.68)	6.89 (2.07)	7.57 (2.10)
NC task-F	6.67 (2.55)	5.63 (1.99)	6.80 (2.02)	8.43 (3.96)
NC task-C	6.46 (2.45)	5.06 (2.45)	7.00 (2.10)	7.79 (2.23)
Implicit comprehension				
NC task-R	5.71 (2.34)	5.03 (2.56)	5.70 (2.09)	7.29 (1.98)
NC task-F	5.60 (2.94)	4.23 (1.83)	5.89 (1.91)	7.57 (5.52)
NC task-C	4.89 (2.40)	4.00 (2.27)	5.00 (2.32)	6.50 (2.21)

Table 7: Task Performance by Book and Reading Ability

NC TASK VARIABLES	NONREADER (N = 70)	READER (N = 21)
Picture walk		
NC task-R	6.97 (1.86)	7.71 (2.28)
NC task-F	7.06 (2.29)	7.95 (1.75)
NC task-C	6.94 (2.32)	7.52 (1.99)
Retelling		
NC task-R	2.81 (1.94)	3.86 (1.49)
NC task-F	2.67 (1.68)	4.10 (1.67)
NC task-C	2.61 (1.77)	4.24 (1.64)
Total comprehension		
NC task-R	11.29 (4.08)	15.14 (2.43)
NC task-F	11.25 (3.40)	13.95 (3.01)
NC task-C	10.10 (4.04)	15.52 (2.29)
Explicit comprehension		
NC task-R	5.91 (2.49)	7.95 (1.60)
NC task-F	6.38 (2.73)	7.62 (1.50)
NC task-C	5.81 (2.37)	8.62 (1.16)
Implicit comprehension		
NC task-R	5.27 (2.32)	7.19 (1.75)
NC task-F	5.38 (3.17)	6.33 (1.93)
NC task-C	4.29 (2.25)	6.90 (1.73)

Inter-task correlations by outcome were also examined to ascertain whether similar scores were obtained by the same students across picture books. The following significant positive correlations emerged for each Picture Walk, Retelling, and Total Prompted Comprehension measure across NC Task versions:

- Picture Walk: $r = .62, p < .01$ between NC task-R and NC task-F;
  $r = .53, p < .01$ between NC task-R and NC task-C;
  $r = .69, p < .01$ between NC task-F and NC task-C
- Retelling: $r = .64, p < .01$ between NC task-R and NC task-F;
  $r = .70, p < .01$ between NC task-R and NC task-C;
  $r = .61, p < .01$ between NC task-F and NC task-C
- Total comprehension: $r = .63, p < .01$ for NC task-R and NC task-F;
  $r = .62, p < .01$ for NC task-R and NC task-C;
  $r = .58, p < .01$ for NC task-F and NC task-C
- These high correlations show that children were consistent on each dependent variable across all versions of the NC task.

Picture Walk. Picture Walk items were examined individually in order to compare the breakdown of 0-1-2 points awarded for each NC task version.

These detailed frequency analyses reveal similar patterns in Picture Walk behaviors for all three picture books. As the Picture Walk behavior increased

Table 8: Intra-Task Correlation Matrix for NC Task-R, NC Task-F, and NC Task-F

	PICTURE WALK	RETELLING	TOTAL COMP	EXPLICIT COMP	IMPLICIT COMP
Picture walk	—				
A. NC task-R		.23* (A)	.20 (A)	.19 (A)	.18 (A)
B. NC task-F		.43† (B)	.43† (B)	.36† (B)	.31† (B)
C. NC task-C		.36† (C)	.37† (C)	.33† (C)	.34† (C)
Retelling		—			
A. NC task-R			.47† (A)	.49† (A)	.33† (A)
B. NC task-F			.46† (B)	.43† (B)	.32† (B)
C. NC task-C			.61† (C)	.55† (C)	.54† (C)
Total comp			—		
A. NC task-R				.88† (A)	.87† (A)
B. NC task-F				.70† (B)	.51† (B)
C. NC task-C				.90† (C)	.89† (C)
Explicit comp				—	
A. NC task-R					.56† (A)
B. NC task-F					.72† (B)
C. NC task-C					.61† (C)
Implicit comp					—
A. NC task-R					
B. NC task-F					
C. NC task-C					

* $p < .05$
† $p < .01$

in complexity, greater percentages of students received zero points for the behavior. For all three books, approximately 1–2% of students received zero points for Book Handling, whereas approximately 59–69% of students received zero points for comprehension strategies. Conversely, fewer students were awarded two points for the more complex behaviors. Approximately 85–89% of students received two points for Book Handling Skills and roughly 11–17% scored 2 points for comprehension strategies. Tables 6 and 7 show Picture Walk scores for children on all three NC Task versions as a function of both grade level and reading ability. ANOVAs by grade level revealed no significant differences for either NC task-R or NC task-C. However, Picture Walk scores did significantly increase by grade for NC task-F, $F(2, 87) = 3.10$, $p < .05$. Picture Walk scores for nonreaders as compared to readers exhibited similar trends, as is demonstrated by the ANOVAs, which show no significant differences based on reading status for all three task versions. Nor did the tasks result in any significant differences by gender or ethnicity.

Retelling. Retelling scores were also similar across NC task versions, with students recalling approximately three major story elements from each of the three picture books. As shown in Table 6, children demonstrated consis-

tent age-related differences across NC task versions. Kindergartners scored approximately two points for each picture book retelling, first graders received roughly three points for their retellings, and second graders recalled approximately four to five main story elements. These grade differences were significant for NC task-R, $F(2, 88) = 6.59$, $p < .01$; NC task-F, $F(2, 87) = 8.79$, $p < .001$; and NC task-C, $F(2, 88) = 9.71$, $p < .001$. Readers and nonreaders scored similarly on all picture books. All nonreaders included approximately three elements in their retellings regardless of NC task version, whereas readers included approximately four main elements in their retellings. Mean retelling scores were significantly different by reading ability for NC task-R, $F(1, 89) = 5.16$, $p < .05$; NC task-F, $F(1, 88) = 11.70$, $p < .001$; and NC task-C, $F(1, 89) = 14.02$, $p < .001$. ANOVAs did not reveal any gender or ethnic differences for any of the picture books.

Prompted Comprehension. The ten prompted comprehension items were examined individually in order to compare the breakdown of 0-1-2 points awarded for each NC task version. These frequency analyses revealed similar patterns in narrative comprehension levels for all three picture books. For all picture books, the highest percentage of students received 0 comprehension points for the setting, prediction, and theme questions. For NC task-C only, many students (45.1%) could not infer characters' feelings. Furthermore, the majority of students received 1 point for Initiating Event and Dialogue, and for most of the other comprehension questions one-quarter to one-half of the children provided one-point answers. Patterns were also similar across picture books for receiving 2 points. The highest percentages of children scored twos for character, causal inference, problem, and outcome resolution questions, whereas fewer students received twos on the initiating event, prediction, theme, dialogue, and feeling questions. These scores indicate that many children focus on events on a single page rather than the narrative level of integration.

Examination of the Prompted Comprehension, Explicit Comprehension, and Implicit Comprehension distributions by grade level indicated similar trends across picture books (see Table 6). Significant age-related increases emerged for these three measures on each picture book. Higher-grade students received significantly more points than lower-grade students on Total Prompted Comprehension for all three NC task versions: NC task-R, $F(2, 88) = 7.18$, $p < .001$; NC task-F, $F(2, 87) = 9.17$, $p < .001$; NC task-C, $F(2, 88) = 9.47$, $p < .001$. Grade level differences emerged on Explicit Comprehension for NC task-R, $F(2, 88) = 7.18$, $p < .001$; NC task-F, $F(2, 87) = 6.62$; and NC task-C, $F(2, 88) = 9.47$, $p < .001$. Similar trends were revealed for Implicit Comprehension: NC task-R, $F(2, 88) = 4.86$, $p < .01$; NC task-F, $F(2, 88) = 7.57$, $p < .001$; and NC task-C, $F(2, 88) = 5.88$, $p < .01$. Means by reading ability were also comparable across picture books, with nonreaders receiving approximately the same number of points on the three comprehension measures for all three picture books and readers also scoring approximately the same number of points on the three measures across picture books (see Table 7). ANOVAs by reading ability revealed significant effects for each picture book on all three measures. Readers scored significantly more points than nonreaders on Total Prompted Comprehension for the three books: NC task-R, $F(1, 89) = 16.91$, $p < .001$; NC task-F, $F(1, 88) = 10.72$, $p < .01$; and NC task-C, $F(1,89) = 34.40$, $p < .001$. For Explicit Comprehension, readers also scored significantly higher than nonreaders: NC task-R, $F(1, 89) = 12.43$, $p < .001$; NC task-F, $F(1, 88) = 3.95$, $p < .05$; and NC task-C, $F(1, 89)$

= 27.35, $p < .001$. Likewise, readers performed better on Implicit Comprehension than did nonreaders for NC task-R, $F(189) = 12.23$, $p < .001$; and for NC task-C, $F(189) = 24.05$, $p < .001$. Readers did not score significantly higher than nonreaders on Implicit Comprehension for NC task-F. No ethnicity or gender differences were found for any of the comprehension measures on any of the NC task versions.

Discussion

Study 2 demonstrates that the NC task yields remarkably consistent results across three different picture books. All five measures of the NC task revealed the same developmental trends for Prompted Comprehension and very few changes in Picture Walk behaviors. The significant changes with age for Picture Walk for the frog book may be due to that book's finer details. The slightly lower comprehension scores for the carpet book may indicate greater narrative complexity, but the patterns of means and correlations are similar across all three books. Similarity of performance across books also shows that examiners can administer the task with different materials and score children's performance in a reliable manner. Thus, the generalizability of the NC task across picture books is supported and the task appears robust.

Study 2 also revealed similar developmental trends in the data by age and reading ability. The NC task is sensitive to progressive increases in the ability to make inferences and connections among pictures and to construct coherent narrative relations from picture books. This narrative ability is also correlated with improvements in reading ability, which may be a by-product of increases in age or may reflect specific experiences with reading and narrative schemata in children's books. Narrative comprehension appears to be an outside-in process strongly related to top-down processes of reading.

Conclusions

The two studies in this paper provide encouraging evidence about children's narrative understanding of picture books. First, the NC task appears to be developmentally appropriate for four- to eight-year-old children in terms of administrative procedures and diagnostic sensitivity. The Picture Walk behaviors vary less among children in this age range than retelling and comprehension measures, but all can provide useful diagnostic information to teachers. The task can be given in less than 15 minutes and provides children with authentic, enjoyable book experiences. Second, the NC task yields reliable, quantifiable data through standard procedures that are generalizable across picture books. Researchers can pit the uniform procedures of the NC task against other outside-in and inside-out processes in subsequent tests of developmental models of reading factors. The NC task may help in a small way to refute claims about the priority of inside-out factors by providing a viable outside-in assessment of prereading processes. At least, it is a model of the kind of outside-in assessment tasks that are needed.

Third, teachers may adapt the NC task to informal classroom use in order to make their own diagnostic assessments of children's thinking about narratives. The task provides information about children's cognitive understanding of stories independently of decoding abilities and is highly consistent with primary-grade instruction on narrative structure in text. Because the NC task can be adapted to any narrative text, it can be used with common classroom curricular materials that present stories in basal readers, oral stories, or visual media. Since the task does not depend on the decoding of English text, it also can easily be adapted for use with bilingual children or in ESL classrooms. Furthermore, the NC task can be used interchangeably by teachers for both assessment and instruction, a hallmark of an educationally useful and authentic task.

Fourth, the NC task provides the first compelling data to support recent claims that children's understanding of narrative is an important foundation for learning to read (Burns, Griffin, & Snow, 1999; Whitehurst & Lonigan, 1998). We believe that narrative competence may be a general feature of children's thinking that is critical for early literacy and cognitive development. It is pervasive in children's language, play, and thinking (Yussen & Ozcan, 1996) and is supported by parents and teachers alike in their normal practices. Narrative competence can be expanded beyond the NC task in this paper in several ways. We plan to extend the present studies by examining how well narrative comprehension predicts children's comprehension on literacy tasks over time. Furthermore, narrative productions in children's language, play, and writing may provide indices of the development and quality of narrative thinking. We imagine that assessments of children's narrative comprehension may show developmental advances in their narrative productions, in print at least, but we expect that even two- to four-year-old children exhibit narrative thinking with pictures, objects, and language. The idea that these experiences may influence literacy development is certainly a viable hypothesis about the content and quality of parent-child joint bookreading and demands further exploration.

We began this paper with a discussion of the bottom-up versus top-down controversy of early reading. The compromise solution to this polarized debate has been a call for "balanced instruction" in the classroom. However, balanced instruction is difficult to achieve if the weights at the ends of the scale are unequal. It is our contention that part of the historic imbalance between outside-in and inside-out factors in early reading is due to disparities in the quantity of assessment tools, the specificity of processes, and the ease of measurement. Balanced instruction depends on balanced theories and balanced assessments; it will not be possible to achieve instructional equilibrium until the inside-out and outside-in factors influencing instructional practices are more nearly equal. It is hoped that tools such as the NC task can begin to provide better balance among the factors that influence early reading.

References

Adams, M. A. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: The MIT Press.

Bruner, J. (1986). *Actual minds, possible worlds*. Cambridge, MA: Harvard University Press.

Burns, M. S., Griffin, P., & Snow, C. E. (1999). *Starting out right*. Washington, DC: National Academy Press.

Chall, J. S. (1967). *Learning to read: The great debate*. New York: McGraw-Hill.

Clay, M. (1985). *Early detection of reading abilities*. (3rd ed.). Portsmouth, NH: Heinemann.

Edmondson, J. (1998). America Reads: Doing battle. *Language Arts, 76*(2), 154-162.

Flesch, R. (1955). *Why Johnny can't read*. New York: Harper and Row.

Fountas, I. C., & Pinnell, G. S. (1996). *Guided reading*. Portsmouth, NH: Heinemann.

Goodman, Y. (1986). Children coming to know literacy. In W. H. Teale & E. Sulzby (Eds.), *Emergent literacy: Writing and reading* (pp. 1-14). Norwood, NJ: Alex Publishing Corporation.

Leslie, L., & Caldwell, J. (1995). *The Qualitative Reading Inventory*. Glenview, IL: Scott Foresman.

Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher, 20*(8), 15-21.

Michigan Department of Education Early Literacy Committee. (1998). *Michigan Literacy Progress Profile*. Lansing, MI: Department of Education.

Morrow, L. M. (1990). Assessing children's understanding of story through their construction and reconstruction of narrative. In L. M. Morrow & J. K. Smith (Eds.), *Assessment for instruction in early literacy* (pp. 110-134). Englewood Cliffs, NJ: Prentice Hall.

Reading Excellence Act. HR2614, 1998.

Snow, C. E., & Ninio, A. (1986). The contracts of literacy: What children learn from learning to read books. In W. H. Teale & E. Sulzby (Eds.), *Emergent literacy: Writing and reading* (pp. 116-138). Norwood, NJ: Ablex Publishing Corporation.

Stallman, A. C., & Pearson, P. D. (1990). Formal measures of early literacy. In M. A. Smith (Ed.), *Assessment for instruction in early literacy* (pp. 7-44).

Sulzby, E. (1985). Children's emergent reading of favorite storybooks: A developmental study. *Reading Research Quarterly, 20*(4), 458-481.

van Kraayenoord, C. E., & Paris, S. (1996). Story construction from a picture book: An assessment activity for young learners. *Early Childhood Research Quarterly, 11*, 41-61.

Whitehurst, G. J., & Lonigan, C. J. (1998). Child development and emergent literacy. *Child Development, 69*(3), 848-872.

Yussen, S. R., & Ozcan, N. (1996). The development of knowledge about narratives. *Issues in Education, 2*(1), 1-68.

Appendix A: NC Task Picture Walk

"We're going to look through this book together, and as we go through it I want you to tell me whatever you are thinking about the pictures or the story."

Picture Walk Element	Score Description	Score
1. **Book Handling Skills:** Orients book correctly, has sense of appropriate viewing speed and order, where viewing errors include skipping pages, speeding through pages, etc.	Incorrectly handles book and makes more than 2 viewing errors	0
	Makes 1-2 viewing errors (i.e., skips pages)	1
	Handles book appropriately, making no viewing errors	2
2. **Engagement:** Behavioral and emotional involvement during picture walk, as judged by attention, interest in book, affect, and effort.	Displays off-task behavior or negative comments	0
	Displays quiet, sustained behavior	1
	Shows several examples of attention, affect, interest, or effort (i.e., spontaneous comments)	2
3. **Picture Comments:** Discrete comments about a picture, which can include descriptions of objects, characters, emotions, actions, and opinions as well as character vocalizations.	Makes no picture comments	0
	Makes 1 picture comment or verbalization	1
	Makes 2 or more comments or verbalizations about specific pictures	2
4. **Storytelling Comments:** Makes comments that go across pictures which demonstrate an understanding that the pictures tell a coherent story—can include narration, dialogue, using book language and storytelling voice.	Makes no storytelling comments	0
	Provides storytelling elements, but not consistently	1
	Through narration or dialogue, connects story events and presents a coherent story line	2
5. **Comprehension Strategies:** Displays vocalizations or behaviors which show attempts at comprehension, such as self-corrects, looks back/ahead in book, asks questions for understanding, makes predictions about story.	Demonstrates no comprehension strategies	0
	Exhibits 1 instance of comprehension strategies	1
	Demonstrates comprehension strategies at least 2 or more times.	2

Appendix B: NC Task Prompted Comprehension Questions

Explicit Questions

1. [Book Closed, Characters]: Who are the characters in this story? (replacement words: people, animals)
2. [Book Closed, Setting]: Where does this story happen? (replacement words: setting, take place)
3. [Pg.10, Initiating Event]: Tell me what happens at this point in the story. Why is this an important part of the story?
4. [Pg.12, Problem]: If you were telling someone this story, what would you say is going on now? Why did this happen?
5. [Pg.18, Outcome Resolution]: What happened here? Why does this happen?

Implicit Questions

1. [Pg.6, Feelings]: Tell me what the people are feeling in this picture. Why do you think so?
2. [Pg.8, Causal Inference]: Why did the family get the robot?
3. [Pg.16, Dialogue]: What do you think the people would be saying here? Why would they be saying that?
4. [Pg.18, Prediction]: This is the last picture in the story. What do you think happens next? Why do you think so?
5. [Book Closed, Theme]: In thinking about everything that you learned after reading this book, if you knew that your friend's dad was bringing home a robot for his family, what would you tell the dad to help him so that the same thing that happened in this story doesn't happen to him? Why would you tell him that? (replacement words: advice, warn)

Appendix C: NC Task Prompted Comprehension Scoring

The purpose of this appendix is to provide a clear set of guidelines that describes what we did and what others should do if they use this task.

Rubrics for Scoring the Prompted Comprehension Questions

Explicit Information

Characters:

- 2 points = response indicates that characters are a family and a robot
- 1 point = response contains at least 2 of the story's characters
- 0 points = response provides only 1 character or answer is inappropriate

Setting

- 2 points = responses indicate an understanding of multiple settings
- 1 point = response provides only one setting
- 0 points = response is not an appropriate setting

Initiating Event

- 2 points = response identifies the initiating event and links it with other relevant story information: e.g., with the problem
- 1 point = response identifies the story element, in this case the initiating event
- 0 points = response fails to identify the initiating event

Problem

- 2 points = response identifies the problem and links it with other relevant story information: e.g., with the initiating action
- 1 point = response identifies the story element, in this case the problem
- 0 points = response fails to identify the problem

Outcome Resolution

- 2 points = response identifies the outcome resolution and links it with other relevant story information: e.g., the problem or the initiating action
- 1 point = response identifies the story element, in this case the initiating action

- 0 points = response fails to identify the outcome resolution

Implicit Information

Feelings

- 2 points = response indicates the inference of appropriate character feelings and connects the feelings to other pages or events
- 1 point = response indicates the inference of appropriate character feelings
- 0 points = response is not an appropriate inference of character feelings

Causal Inference

- 2 points = response is an appropriate inference that is explained by using events from multiple pages
- 1 point = response is an appropriate inference that is derived at the page level
- 0 points = response fails to include an appropriate causal inference

Dialogue

- 2 points = response indicates the inference of appropriate character dialogue and connects the dialogue to other pages or events
- 1 point = response indicates the inference of appropriate character dialogue
- 0 points = response does not concern character dialogue or is not appropriate

Prediction

- 2 points = response represents a prediction that used previous action or pages from the story
- 1 point = response indicates a prediction that could be made based only on the last picture of the story
- 0 points = response does not contain an appropriate prediction

Theme

- 2 points = response indicates the incorporation of multiple events in order to create a narrative-level theme
- 1 point = response is a simple theme that uses information from one aspect of the story
- 0 points = response does not indicate an understanding of any theme

Examples

The following are examples of 0-, 1-, and 2-point responses to the initiating event question (A) and the prediction question (B). On the page for which the child is asked to describe the initiating event, there is a picture of a girl pulling out the wires of the robot, which leads to the robot's becoming wild and ruining the house, the problem of the story. On the final page, from which the child is asked to infer a prediction, the father is fixing the robot, suggesting that the robot will be able to clean the house as it did when it was new.

0 Points: Fails to Identify Element or Make Inference

A. "She's cleaning the robot. [This is important] because it's always nice to get cleaned, isn't it?"

B. "It's just the end. [I know this] because I don't see any more pages below it"

1 Point: Picture-Level Responses

A. "The little girl is undoing all the cords and she's going to tie them into a bow so it looks like a girl. [This is important] maybe because she wants him to look more like a girl."

B. "It works again. [I know this] because they're fixing it."

2 Points: Narrative-Level Responses

A. "The girl pulls out all of the wires. [This is important] because if we didn't know this, we wouldn't know why it was acting up."

B. "Maybe the machine tries to go away but it gets caught by them. [I know this] because he's getting tired of doing all the chores."

Appendix D: QRI-II

QRI-II measures: The QRI-II includes graded word lists that are used to estimate children's initial reading levels; graded informational and narrative passages from preprimer through junior high levels; procedures for eliciting and scoring retellings; and comprehension questions corresponding to each passage about explicit and implicit text information.

Graded Word Lists: Children were first given a word list at their current grade level and then a second list at the next grade level higher or lower, depending on their performance on the first word list. Certain students received more than two word lists if their reading level was far enough away from the initial grade level list that more lists were required to arrive at the child's actual level. Children who could not read the preprimer list did not receive a second, more difficult list.

Oral Reading: The first passage given to the child was selected based on 90–100% correct on the equivalent grade-level word list, and the second passage was generally one grade-level higher, unless the student read the first passage below the frustration level (below 90%), in which case the child received the next lower level passage. Children's oral reading was measured through miscue analysis, in which the tape-recorded oral readings were scored for omitted, substituted, or inserted words in addition to self-corrected miscues. Percent accuracy was based on total miscues; percent acceptability was based only on meaning-changing miscues; and percent of self-corrections was calculated as both a percent of total miscues and a percent of only meaning-changing miscues.

Appendix E: MLPP

1. Six scores were obtained from the three MLPP activities: one for each of the three tasks, and three additional subscores derived from the Phonemic Awareness task.

2. The book used to assess Concepts About Print possessed the features designated by the MLPP guidelines, including at least one example of: a) print and illustration on a single page, b) multiple lines of text on a single page, and c) a variety of punctuation marks. The book that we used was *Wake Up, Sun*, written by David L. Harrison and illustrated by Hans Wilhelm. This book was labeled a "Step 1" book in the series "Step into Reading." Locations were marked in the book where the specific Concepts of Print questions would be asked (including the front, back, and individual inside pages).

3. Examples of each of the Concepts of Print categories: Book Concepts (e.g., front cover, back cover, title); Reading Concepts (print carries message, one-to-one match); Directionality (e.g., beginning of text, left to right and top to bottom, return sweep); Concept of Word (e.g., first word, last word, word); Concept of Letter (e.g., first letter in word, last letter in word, one letter/two letters, letter names, capital letter, small letter); and Punctuation Marks (e.g., period, question, exclamation, quotation, comma).

4. In the rhyming section of the Phonemic Awareness task, the child was provided with a definition of rhyming and a few practice words. They were then given a series of eight words and asked to provide a word that rhymed with each item on the list (e.g., "Tell me a word that rhymes with 'bat'"). If the child provided a nonsense rhyming word, (s)he was asked "can you tell me another word that is a real word?" The child received 1 point for each correct rhyme; total Rhyme scores could range from 0–8 points.

   For the Phoneme Blending section, the child received a definition of what it means to "put sounds together" and was allowed to practice with a few examples. Then they were presented with a series of sounds and asked to blend the sounds together into words. For example, the experimenter would say, "What word would I have if I put together the sounds /t/ /a/ /p/?" The child was asked to blend together eight examples of three-sound words, and 1 point was earned for each correct word blend. Phoneme Blending scores range from 0–8 points.

   The Segmentation section assesses whether children can discern sounds in words clearly enough to reproduce them in print. The child first listened to an explanation of what it means to "stretch out a word... by thinking about how many sounds you hear", and then received several practice items. They were asked "what are the sounds?" for a list of eight words and received one point for each correct response. Again, scores for Segmentation range from 0–8 points.

   If the child completely missed three consecutive items on any of the three phonemic awareness tasks and appeared confused about what was being assessed, the task was discontinued and the child was advanced to the next task in order to ensure that their experience was a positive one.

5. The following is the two-sentence "story": "I see a white cat in the sun. It is looking for some big toys." This is the kindergarten-level story, which all children received because they had all been identified as nonreaders. These sentences were read to the child at a normal pace and then dictated again very slowly so that the child could write down the words as they were heard.

About CIERA

The Center for the Improvement of Early Reading Achievement (CIERA) is the national center for research on early reading and represents a consortium of educators in five universities (University of Michigan, University of Virginia, and Michigan State University with University of Southern California and University of Minnesota), teacher educators, teachers, publishers of texts, tests, and technology, professional organizations, and schools and school districts across the United States. CIERA is supported under the Educational Research and Development Centers Program, PR/Award Number R305R70004, as administered by the Office of Educational Research and Improvement, U.S. Department of Education.

Mission. CIERA's mission is to improve the reading achievement of America's children by generating and disseminating theoretical, empirical, and practical solutions to persistent problems in the learning and teaching of beginning reading.

CIERA Research Model

The model that underlies CIERA's efforts acknowledges many influences on children's reading acquisition. The multiple influences on children's early reading acquisition can be represented in three successive layers, each yielding an area of inquiry of the CIERA scope of work. These three areas of inquiry each present a set of persistent problems in the learning and teaching of beginning reading:

CIERA INQUIRY 1
Readers and Texts

Characteristics of readers and texts and their relationship to early reading achievement. What are the characteristics of readers and texts that have the greatest influence on early success in reading? How can children's existing knowledge and classroom environments enhance the factors that make for success?

CIERA INQUIRY 2
Home and School

Home and school effects on early reading achievment. How do the contexts of homes, communities, classrooms, and schools support high levels of reading achievement among primary-level children? How can these contexts be enhanced to ensure high levels of reading achievement for all children?

CIERA INQUIRY 3
Policy and Profession

Policy and professional effects on early reading achievement. How can new teachers be initiated into the profession and experienced teachers be provided with the knowledge and dispositions to teach young children to read well? How do policies at all levels support or detract from providing all children with access to high levels of reading instruction?

www.ciera.org

CIERA
Center for the Improvement of Early Reading Achievement

CIERA is a collaboration of:
University of Michigan
University of Virginia
Michigan State University
with:
University of Minnesota
University of Southern California

University of Michigan School of Education
610 E University Av, Rm 1600 SEB
Ann Arbor, MI 48109-1259
734.647.6940 voice
734.763.1229 fax
ciera@umich.edu

www.ciera.org